Generative Adversarial Networks and its applications

Takaya KOIZUMI

Mathematical Science, B4

Applied Mathematics and Physics informal seminor, 2nd

Contents

Contents

Generarive Adversarial Networks

敵対的生成ネットワーク (Generative Adversarial Networks, GAN) は Goodfellow らによって生み出された, 2 つのニューラルネットワークを用いる教師なし学習の一種である. 現在も様々な亜種が多く考案されており, 近年様々なアプリに応用されている.

GAN のアーキテクチャ

Definition (Generative Adversarial Networks[1])

ML 空間 $(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を以下のように定義する.
$\mathcal{X}$ は $\mathbb{R}^d$ のコンパクト部分集合, $\mathcal{Y} = [0, 1]$,

$$\mathcal{H}_G = \{G : \mathcal{Z} \to \mathcal{X} \mid G \text{ はニューラルネット }\},$$
$$\mathcal{H}_D = \{D : \mathcal{X} \to \mathcal{Y} \mid D \text{ はニューラルネット }\},$$
$$\mathcal{L}_{\mathbb{D}}(G, D) = \mathbb{E}_{x \sim \mathbb{P}_{data}}[\log D(x)] + \mathbb{E}_{x' \sim \mathbb{P}_G}[\log(1 - D(x'))],$$

ここで, $\mathcal{Z}$ は潜在空間と呼ばれる $\mathbb{R}^d$ の部分空間である. また, $\mathbb{P}_G$ は $G \in \mathcal{H}_G$ と確率分布 $\mathbb{P}_Z$(一様分布や正規分布) に従う確率変数 $Z : \Omega \to \mathcal{Z}$(ノイズ) に対し, $G(Z)$ が従う確率分布である. この時,

$$\arg\min_{G \in \mathcal{H}_G} \arg\max_{D \in \mathcal{H}_D} \mathcal{L}_{\mathbb{D}}(G, D).$$

を求める問題を敵対的生成ネットワーク *(GAN)* という.

Generator と Discriminator

Definition (Generator と Discriminator)

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を *GAN* とする. $G \in \mathcal{H}_G$ を生成器 *(Generator)* と呼び, $D \in \mathcal{H}_D$ を判別器 *(Discriminator)* と呼ぶ. また, 経験損失関数 $\mathcal{L}_{\mathbb{D}}$ を *Adversarial loss* という. さらに, $G \in \mathcal{H}_G$ に対して $D_G^* \in \mathcal{H}_D$ が,

$$\forall D \in \mathcal{H}_D, \mathcal{L}_{\mathbb{D}}(G, D_G^*) \geq \mathcal{L}_{\mathbb{D}}(G, D)$$

を満たす時, D_G^* は G に関しての最適 *Discriminator* であるという.

Notation 2.

今後, p_{data} は $\mathbb{P}_{data}$ の確率密度関数であり, p_G は $\mathbb{P}_G$ の確率密度関数を表すものとする.

Contents

GAN のやりたいこと

今まで敵対的生成ネットワークの構造を話してきたが, ここでは GAN の成し遂げたいことについて解説する. GAN の目的は「データの確率密度関数 p_{data} をニューラルネット G を用いて近似し, あたえられたデータにそっくりなデータ $G(z)$ を生成すること」である.
これを実現するために GAN では G と D のいたちごっこ (min-max ゲーム) を行っている.

GAN のいたちごっこ

- G は偽物を作成する.
- D は与えられたデータが本物かどうかを判定する (本物である確率を返す).

すなわち, G と D を敵対させて学習させ, データ $G(z)$ を生成する手法が GAN である.

いたちごっこの実現と Adversarial loss

GAN でいたちごっこを実現させる要は GAN の経験損失関数 (Adversarial loss),

$$\begin{aligned}\mathcal{L}_{\mathbb{D}}(G, D) &= \mathbb{E}_{x\sim\mathbb{P}_{data}}[\log D(x)] + \mathbb{E}_{x'\sim\mathbb{P}_G}[\log(1 - D(x'))] \\ &= \mathbb{E}_{x\sim\mathbb{P}_{data}}[\log D(x)] + \mathbb{E}_{z\sim\mathbb{P}_Z}[\log(1 - D(G(z)))]\end{aligned}$$

である. x は本物のデータ, $G(z)$ は偽物のデータであるから, $D(x)$, $1 - D(G(z))$ が大きくなるように学習させると $D(y)$ はデータ y が本物である確率に近づく. 逆に, $D(y)$ はデータ y が本物である確率であるから, $1 - D(G(z))$ が小さくなるように学習させると $G(z)$ は本物のデータに近づく. すなわち, $\mathcal{L}_{\mathbb{D}}(G, D)$ を $G \in \mathcal{H}_G$ に関して最小化, $D \in \mathcal{H}_D$ に関して最大化させることがいたちごっこを実現していることがわかる.

Contents

最適 Discriminator

ここからは実際にいたちごっこの結果, p_G が p_{data} と等しくなることを示す.

Proposition (Optimal Discriminator[1])

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を *GAN* とし, $G \in \mathcal{H}_G$ とする. この時, 最適 *Discriminator* は以下で与えられる.

$$D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_G(x)}.$$

命題の証明

Proof.

期待値の計算から

$$\begin{aligned}\mathcal{L}_{\mathbb{D}}(G,D) &= \int_{\mathcal{X}} p_{data}(x)\log(D(x))dx + \int_{\mathcal{X}} p_G(x)\log(1-D(x))dx \\ &= \int_{\mathcal{X}} p_{data}(x)\log(D(x)) + p_G(x)\log(1-D(x))dx\end{aligned}$$

を得る. ここで, $(a,b) \in \mathbb{R}^2 - (0,0)$ に対して, $(0,1)$ 上の関数 $F(y) = a\log(y) + b\log(1-y)$ が,

$$\arg\max_{y\in(0,1)} F(y) = \frac{a}{a+b}$$

を満たすことを用いれば主張が従う. □

virtual training criterion

先の命題より $\mathcal{H}_G$ 上の関数 $C : \mathcal{H}_G \to \mathbb{R}$ が

$$C(G) = \mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log \frac{p_{data}(x)}{p_{data}(x)+p_G(x)}\right] + \mathbb{E}_{x'\sim\mathbb{P}_G}\left[\log \frac{p_G(x')}{p_{data}(x')+p_G(x')}\right]$$

で定まる. 関数 C を仮想訓練基準 (virtual training criterion) と呼ぶ.

Theorem (GAN の最小性 [1])

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を *GAN*, $C : \mathcal{H}_G \to \mathbb{R}$ を仮想訓練基準とする. C が最小値 $-\log 4$ を取るための必要十分条件は $\mathbb{P}_{data} = \mathbb{P}_G$ となることである.

証明

Proof.

$\mathbb{P}_{data} = \mathbb{P}_G$ とする．この時，

$$
\begin{aligned}
C(G) &= \mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log\frac{p_{data}(x)}{p_{data}(x)+p_G(x)}\right] + \mathbb{E}_{x'\sim\mathbb{P}_G}\left[\log\frac{p_G(x')}{p_{data}(x')+p_G(x')}\right] \\
&= \mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log\frac{1}{2}\right] + \mathbb{E}_{x'\sim\mathbb{P}_G}\left[\log\frac{1}{2}\right] \\
&= -\log 2 - \log 2 = -\log 4
\end{aligned}
$$

であるから，C は常に最小値 $-\log 4$ をとる．逆に C の最小値が $-\log 4$ であるとする．JSD を Jensen-Shannon Divergence とし，C を変形すると

$$
C(G) = -\log 4 + JSD\left(\mathbb{P}_{data}\|\mathbb{P}_G\right)
$$

となる．ここで $JSD\left(\mathbb{P}_{data}\|\mathbb{P}_G\right) = 0$ と $\mathbb{P}_{data} = \mathbb{P}_G$ は同値だから $\mathbb{P}_{data} = \mathbb{P}_G$ である． □

Contents

GAN の学習不安定性

前節で GAN の学習について述べたが, 述べた枠組みでは Discriminator が最適化に近づけは近づくほど, GAN の学習が不安定 (G の学習勾配が 0 になって学習が進まなくなる) ことを示す. この subsection では以下の仮定を P とおく.

assumption P

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を GAN とする. この時, $\mathcal{X}$ のコンパクト部分集合 P, M が存在して,

$$M \cap P = \emptyset, \mathrm{supp}(p_{data}) \subset M \text{ かつ } \mathrm{supp}(p_G) \subset P$$

が成立する.

Perfect Discriminator Theorem

Theorem (The Perfect Discriminator Theorem[3])

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}, \mathcal{L}_{\mathbb{D}})$ を *GAN* とする. 仮定 P が成り立つならば, 以下の性質を満たす *smooth* な最適 *Discriminator* D^* が存在する.

1. $\mathbb{P}_{data}(D^* = 1) = 1$ かつ $\mathbb{P}_G(D^* = 0) = 1$
2. $\forall x \in M \cup P, \nabla_x D^*(x) = 0.$

なお, 最大値の一意性より, $D^*_G = D^*$ であるから, 以下この命題の D^* を D^*_G と表記する.

証明

Proof.

$P \cap M = \emptyset$ だから, $\delta = d(P, M)$ とすれば $\delta > 0$ である. ここで,

$$\hat{M} = \{x \in \mathcal{X} \mid d(x, M) \leq \delta/3\}, \quad \hat{P} = \{x \in \mathcal{X} \mid d(x, P) \leq \delta/3\}$$

と定義する. M, P がコンパクトであることと, δ の定義より $\hat{M}$, $\hat{P}$ は共にコンパクトであり, $\hat{M} \cap \hat{P} = \emptyset$ である. したがって, Urysohn's smooth lemma より

$$\exists D^* \in \mathcal{H}_D : \text{smooth s.t. } D^*|_{\hat{M}} = 1 \text{ かつ } D^*|_{\hat{P}} = 0$$

が成立する. 任意の $x \in \mathrm{supp}(p_{data})$ に対して, $D^*(x) = 1$ だから, $\log D_G^*(x) = 0$ である. また, 任意の $x \in \mathrm{supp}(p_G)$ に対して, $D^*(x) = 0$ だから, $\log(1 - D_G^*(x)) = 0$ である. これより D^* が最適 Discriminator であること, 及び 1. が従う. また D^* は $M \cup P$ 上で定値写像だから 2. が成立する. □

Vanishing gradient on the Generator

以下, $D \in \mathcal{H}_D$ のノルムを以下で定義する.

$$\|D\| = \sup_{x \in \mathcal{X}} |D(x)| + \|\nabla_x D(x)\|_2$$

Theorem (Vanishing gradient on the Generator)

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}, \mathcal{L}_{\mathbb{D}})$ を *GAN* とする. 仮定 P および, $\varepsilon > 0$ を任意にとる. ある $M > 0$ が存在して

$$\forall D \in \mathcal{H}_D, \|D_G - D_G^*\| < \varepsilon \text{ かつ } \mathbb{E}_{z \sim \mathbb{P}_Z}[\|J_\theta G_\theta(z)\|_2^2] \leq M^2$$

が成立するとする. この時,

$$\forall D \in \mathcal{H}_D, \|\nabla_\theta \mathbb{E}_{z \sim \mathbb{P}_Z}[\log(1 - D(G_\theta(z)))]\|_2 < M \frac{\varepsilon}{1 - \varepsilon}$$

が成立する *(証明は [3] Thm 2.4. をみよ).*

学習不安定性

Corollary

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}, \mathcal{L}_{\mathbb{D}})$ を *GAN* とする．仮定 P および，$\varepsilon > 0$ を任意にとる．ある $M > 0$ が存在して

$$\forall D \in \mathcal{H}_D, \|D_G - D_G^*\| < \varepsilon \text{ かつ } \mathbb{E}_{z\sim\mathbb{P}_Z}[\|J_\theta G_\theta(z)\|_2^2] \leq M^2$$

が成立するとする．この時，

$$\lim_{D\to D_G^*} \nabla_\theta \mathbb{E}_{z\sim\mathbb{P}_Z}[\log(1 - D(G_\theta(z)))] = 0$$

が成立する．

Contents

学習の安定化

前節では GAN の学習が不安定になることを述べた．ここでは，Miyato らによって開発された行列のスペクトル (最大特異値) を用いて GAN の学習を安定化させる手法である Spectral Normalization について述べる．

Notation

$n \times m$ 行列 A に対して，A の作用素ノルムを $\|A\|_{op}$ と表す．また，Lipschitz 連続関数 $f : \mathcal{X} \to \mathbb{R}$ に対して，Lipschitz ノルムを $\|f\|_{Lip}$ と表す．

Spectral Normalization

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を GAN とする．f を D から最終層の活性化関数 $\mathcal{A}$ を省いたものとする (すなわち $D = \mathcal{A} \circ f$.)

Proposition

f の各層の活性化関数の *Lipschitz* ノルムが *1* であるとする．この時,

$$\|f\|_{Lip} \leq \prod_{k=1}^{K+1} \|W_k\|$$

が成立する．ここで $g_k = \eta(W_k x + b_k)$ とした時,
$f(x) = W_{K+1}^{\top} g_K \circ \cdots \circ g_1(x) + b$ とする．

証明

Proof.

Lipschitz ノルムの性質と $\|g_k\|_{Lip} = \|W_k\|_{op}$ より,

$$\begin{aligned}\|f\|_{Lip} &\leq \prod_{k=1}^{K} \|g_k\|_{Lip} \\ &= \prod_{k=1}^{K} \|\eta_k\|_{Lip}\|W_k\|_{op} \\ &= \prod_{k=1}^{K} \|W_k\|_{op}.\end{aligned}$$

□

Spectral Normalization Generative Adversarial Networks

したがって, 各層のパラメータ W_K の各成分を $\|W_k\|_{op}$ で割れば, $\|f\|_{Lip} \leq 1$ とすることができる. この手法を, Spectral Normalization と呼ぶ ($\|A\|_{op}$ は A の最大特異値に等しい.)

Definition (SNGAN[4])

$(\mathcal{X}, \mathcal{Y}, \mathbb{D}, \mathcal{H}_G \times \mathcal{H}_D, \mathcal{L}_{\mathbb{D}})$ を *GAN* の *ML* 空間とする. この時, *Spectral Normalization* を用いて

$$\arg\min_{G \in \mathcal{H}_G} \quad \arg\max_{D \in \mathcal{H}_D, \|f\|_{Lip} \leq 1} \quad \mathcal{L}_{\mathbb{D}}(G, D).$$

を解く問題を *SNGAN* という. ここで, f は D から最終層の活性化関数 $\mathcal{A}$ を省いたものである (すなわち $D = \mathcal{A} \circ f$.)

Contents

Cycle-Consistent Adversarial Networks

Contents

MakeGirlsMoe と Crypko

アニメに出てくるような画像が作れる.

Figure: MakeGirlsMoe (https://make.girls.moe /)

Crypko (https://crypko.ai/beta) を使えはもっとアニメチックなものが作れる (現在開発中.)

Contents

Deepfake

近代社会において Deepfake(GAN を応用した技術) で以下のような犯罪が発生している.
(https://www.nikkei.com/article/DGXMZO64577690S0A001C2CZ8000/, "「ディープフェイク」脅威に　国内初摘発、海外被害も", 日本経済新聞, 2020 年 11 月 27 日午前 1 時頃閲覧)

- 合成ポルノの作成
- 会社の役員の音声複製
- トランプ (元) 大統領のフェイク画像

面白い技術だけどみんなでルールを守ることが大切!.

References

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David, Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, 2014.

[2] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, International Conference on Computer Vision, 2017

[3] Martin Arjovsky and Leon Bottou, Towards Principled Methods for Training Generative Adversarial Networks, International Conference on Learning Representations, 2017.

[4] Takeru Miyato, Toshiki Kataoka, Masanori Koyama and Yuichi Yoshida, Spectral Normalization for Generative Adversarial Networks, International Conference on Learning